Metabolic Reconstruction of Less Characterized Microorganisms: A New Methodology for Reaction Identification from Genome Sequencing Data

Oscar Dias,[1,*] Miguel Rocha,[2] Isabel Rocha,[1] and Eugénio Campos Ferreira[1]

[1] *IBB - Institute for Biotechnology and Bioengineering*
[2] *CCTC - Computer Science Technology Center*

MOTIVATION:

The Genome-Scale Reconstruction of Metabolic Networks encompasses several steps, such as genome annotation, reactions identification and stoichiometry determination, compartmentation, determination of the biomass composition, energy requirements and additional constraints.

According to the IMG system [4] [*a*] there are currently more than 4.000 genomes (4.355 as of July 2009) fully sequenced with more than 500 (535 as of July 2009) being drafted right now. Sequence similarities between genes and genomes can be established using well established algorithms such as BLAST [1] or FASTA [3].

The yeast *Kluyveromyces lactis*, for which the complete genome sequence is available [2], is attracting increasing attention due to its ability to grow on lactose as a sole carbon source.

Whereas the reconstruction of the metabolic network of a given organism is becoming a widespread procedure, starting with the fully sequenced and (partially) annotated genome sequence, there are still many improvements needed in the current methodologies and a clear lack of computational tools for many of the steps.

In this work, a method which is able to collect online similarity information from genome sequencing data, using the BLAST algorithm, and store it on a relational database for later manual curation is proposed. Also, a computational tool implementing the proposed approach was developed. The genome of the yeast *K. lactis* will be used as a case study for this method, providing information for the first stage of the reconstruction of this eukaryote.

CONCLUSIONS:

The outcome of the alignment of *K. lactis* genes sequences to the NCBI database, demonstrated that the developed method provides valuable information for the reconstruction of a preliminary metabolic network.

The identification of several protein coding sequences and subsequent functions assignment, allowed to conclude that the developed framework can be useful for the reconstruction of genome-scale metabolic networks, especially for those organisms that lack information about the proteins encoded in its genome, such as the case study.

[1] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. BASIC LOCAL ALIGNMENT SEARCH TOOL. *JOURNAL OF MOLECULAR BIOLOGY*, 215(3):403–410, OCT 5 1990.

[2] B. Dujon *et al.* Genome evolution in yeasts. *NATURE*, 430(6995):35–44, JUL 1 2004.

[3] D. Lipman and W. Pearson. RAPID AND SENSITIVE PROTEIN SIMILARITY SEARCHES. *SCIENCE*, 227(4693):1435–1441, 1985.

[4] V. M. Markowitz, E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I.-M. A. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *NUCLEIC ACIDS RESEARCH*, 36 (Sp. Iss. SI):D528–D533, JAN 2008.

[*a*] Integrated Microbial Genomes

Keywords: Metabolic Model, *Kluyveromyces lactis*, Genome-Scale Models Reconstruction, BLAST